

# **STATISTICS**

## **INTRODUCTION**

Statistical tools are found useful in progressively increasing number of disciplines. Statistics originated as stratcraft and has grown markedly. It aids individuals as well as organisations. Governments and private enterprises alike increasingly use the statistical techniques. Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Commerce, Biology, Medicine, Industry, planning, education and so on. As on date there is no other human walk of life, where statistics cannot be applied.

## **ORIGIN AND GROWTH OF STATISTICS**

The word 'Statistics' and 'Statistical' are all derived from the Latin word Status, means a political state. The theory of statistics as a distinct branch of scientific method is of comparatively recent growth. Research particularly into the mathematical theory of statistics is rapidly proceeding and fresh discoveries are being made all over the world.

## **MEANING OF STATISTICS**

Statistics are numerical statement of facts in any department of enquiry placed in relation to each other. - A.L. Bowley

Bowley gives another definition for statistics, which states 'statistics may be rightly called the scheme of averages'.

Statistics may be called the science of counting in one of the departments due to Bowley.

According to Croxton and Cowden Statistics may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis.

## **FUNCTIONS OF STATISTICS:**

There are many functions of statistics.

### **Condensation:**

Generally speaking by the word 'to condense', we mean to reduce or to lessen. Condensation is mainly applied at embracing the understanding of a huge mass of data by providing only few observations. If in a particular class in Chennai School, only marks in an examination are given, no purpose will be served. Instead if we are given the average mark in that particular examination, definitely it serves the better purpose. Similarly the range of marks is also

another measure of the data. Thus, Statistical measures help to reduce the complexity of the data and consequently to understand any huge mass of data.

### **Comparison:**

Classification and tabulation are the two methods that are used to condense the data. They help us to compare data collected from different sources. Grand totals, measures of central tendency measures of dispersion, graphs and diagrams, coefficient of correlation etc provide ample scope for comparison. If we have one group of data, we can compare within itself. If the rice production (in Tonnes) in Tanjore district is known, then we can compare one region with another region within the district. Or if the rice production (in Tonnes) of two different districts within Tamilnadu is known, then also a comparative study can be made. As statistics is an aggregate of facts and figures, comparison is always possible and in fact comparison helps us to understand the data in a better way.

### **Forecasting:**

By the word forecasting, we mean to predict or to estimate before hand. Given the data of the last ten years connected to rainfall of a particular district in Tamilnadu, it is possible to predict or forecast the rainfall for the near future. In business also forecasting plays a dominant role in connection with production, sales, profits etc. The analysis of time series and regression analysis plays an important role in forecasting.

### **Estimation:**

One of the main objectives of statistics is drawn inference about a population from the analysis for the sample drawn from that population. The four major branches of statistical inference are

1. Estimation theory
2. Tests of Hypothesis
3. Non Parametric tests
4. Sequential analysis

In estimation theory, we estimate the unknown value of the population parameter based on the sample observations. Suppose we are given a sample of heights of hundred students in a school, based upon the heights of these 100 students, it is possible to estimate the average height of all students in that school.

## **SCOPE OF STATISTICS**

Statistics is not a mere device for collecting numerical data, but as a means of developing sound techniques for their handling, analyzing and drawing valid inferences from them. Statistics is applied in every sphere of human activity – social as well as physical – like Biology, Commerce, Education, Planning, Business Management, Information Technology, etc. It is almost impossible to find a single department of human activity where statistics cannot be applied. We now discuss briefly the applications of statistics in other disciplines.

### **Statistics and Industry:**

Statistics is widely used in many industries. In industries, control charts are widely used to maintain a certain quality level. In production engineering, to find whether the product is conforming to specifications or not, statistical tools, namely inspection plans, control charts, etc., are of extreme importance. In inspection plans we have to resort to some kind of sampling – a very important aspect of Statistics.

### **Statistics and Commerce:**

Statistics are lifeblood of successful commerce. Any businessman cannot afford to either by under stocking or having overstock of his goods. In the beginning he estimates the demand for his goods and then takes steps to adjust with his output or purchases. Thus statistics is indispensable in business and commerce.

As so many multinational companies have invaded into our Indian economy, the size and volume of business is increasing. On one side the stiff competition is increasing whereas on the other side the tastes are changing and new fashions are emerging. In this connection, market survey plays an important role to exhibit the present conditions and to forecast the likely changes in future.

### **Statistics and Agriculture:**

Analysis of variance (ANOVA) is one of the statistical tools developed by Professor R.A. Fisher, plays a prominent role in agriculture experiments. In tests of significance based on small samples, it can be shown that statistics is adequate to test the significant difference between two sample means. In analysis of variance, we are concerned with the testing of equality of several population means. For an example, five fertilizers are applied to five plots each of wheat and the yield of wheat on each of the plots is given. In such a situation, we are interested in finding out whether the effect of these fertilizers on the yield is significantly different or not. In other words, whether the samples are

drawn from the same normal population or not. The answer to this problem is provided by the technique of ANOVA and it is used to test the homogeneity of several population means.

### **Statistics and Economics:**

Statistical methods are useful in measuring numerical changes in complex groups and interpreting collective phenomenon. Nowadays the uses of statistics are abundantly made in any economic study. Both in economic theory and practice, statistical methods play an important role. Alfred Marshall said, “Statistics are the straw only which I like every other economists have to make the bricks”. It may also be noted that statistical data and techniques of statistical tools are immensely useful in solving many economic problems such as wages, prices, production, distribution of income and wealth and so on. Statistical tools like Index numbers, time series Analysis, Estimation theory, Testing Statistical Hypothesis are extensively used in economics.

### **Statistics and Education:**

Statistics is widely used in education. Research has become a common feature in all branches of activities. Statistics is necessary for the formulation of policies to start new course, consideration of facilities available for new courses etc. There are many people engaged in research work to test the past knowledge and evolve new knowledge. These are possible only through statistics.

### **Statistics and Planning:**

Statistics is indispensable in planning. In the modern world, which can be termed as the “world of planning”, almost all the organizations in the government are seeking the help of planning for efficient working, for the formulation of policy decisions and execution of the same. In order to achieve the above goals, the statistical data relating to production, consumption, demand, supply, prices, investments, income expenditure etc and various advanced statistical techniques for processing, analyzing and interpreting such complex data are of importance. In India statistics play an important role in planning, commissioning both at the central and state government levels.

### **Statistics and Medicine:**

In Medical sciences, statistical tools are widely used. In order to test the efficiency of a new drug or medicine, t - test is used or to compare the efficiency of two drugs or two medicines,

ttest for the two samples is used. More and more applications of statistics are at present used in clinical investigation.

### **Statistics and Modern applications:**

Recent developments in the fields of computer technology and information technology have enabled statistics to integrate their models and thus make statistics a part of decision making procedures of many organizations. There are so many software packages available for solving design of experiments, forecasting simulation problems etc. SYSTAT, a software package offers mere scientific and technical graphing options than any other desktop statistics package. SYSTAT supports all types of scientific and technical research in various diversified fields as follows

1. Archeology: Evolution of skull dimensions
2. Epidemiology: Tuberculosis
3. Statistics: Theoretical distributions
4. Manufacturing: Quality improvement
5. Medical research: Clinical investigations.
6. Geology: Estimation of Uranium reserves from ground water

### **LIMITATIONS OF STATISTICS:**

Statistics with all its wide application in every sphere of human activity has its own limitations. Some of them are given below.

**1. Statistics is not suitable to the study of qualitative phenomenon:** Since statistics is basically a science and deals with a set of numerical data, it is applicable to the study of only these subjects of enquiry, which can be expressed in terms of quantitative measurements. As a matter of fact, qualitative phenomenon like honesty, poverty, beauty, intelligence etc, cannot be expressed numerically and any statistical analysis cannot be

directly applied on these qualitative phenomenon's. Nevertheless, statistical techniques may be applied indirectly by first reducing the qualitative expressions to accurate quantitative terms. For example, the intelligence of a group of students can be studied on the basis of their marks in a particular examination.

**2. Statistics does not study individuals:** Statistics does not give any specific importance to the individual items, in fact it deals with an aggregate of objects. Individual items, when they are taken

individually do not constitute any statistical data and do not serve any purpose for any statistical enquiry.

**3. Statistical laws are not exact:** It is well known that mathematical and physical sciences are exact. But statistical laws are not exact and statistical laws are only approximations. Statistical conclusions are not universally true. They are true only on an average.

**Statistics table may be misused:** Statistics must be used only by experts; otherwise, statistical methods are the most dangerous tools on the hands of the inexpert. The use of statistical tools by the inexperienced and untraced persons might lead to wrong conclusions. Statistics can be easily misused by quoting wrong figures of data. As King says aptly 'statistics are like clay of which one can make a God or Devil as one pleases'.

**5. Statistics is only, one of the methods of studying a problem:** Statistical method does not provide complete solution of the problems because problems are to be studied taking the background of the countries culture, philosophy or religion into consideration. Thus the statistical study should be supplemented by other evidences.

### **Collection of Data**

Sources of data are two kinds

Primary Data and Secondary Data

The data which is collected by actual observation or measurement is called primary data. The data which is compiled from the records of others is called secondary data.

**Following are given the methods for collection of such data:**

(i) Direct personal investigation.

(ii) Indirect oral investigation.

iii) Through local correspondents.

(iv) Mailed Questionnaire.

(v) Schedules sent through enumerators.

### **1. Direct personal Investigation:**

ADVERTISEMENTS:

Here, the investigator himself visits the persons those are source of the data and collects necessary information either through interview with the persons concerned or through observation of the data on the spot. This method is suitable where intensive study of any phenomenon is required.

**This method has many merits and demerits which are given as following:**

**Merits:**

1. The data is more reliable as they are obtained directly.

**ADVERTISEMENTS:**

2. Sensitive questions can be avoided.

3. There is chance of more response in it.

4. The questions can be adjusted according to the standard of the sources.

5. There is uniformity in data.

**Demerits:**

1. It is not suitable where the field is very vast and wide.

2. It is very much expensive.

3. It needs a large number of enumerators.

4. It takes long time to collect the data from all the persons.

**ADVERTISEMENTS:**

5. It is purely subjective in nature, therefore the success of investigation depends mostly intelligence, skill, tact, insight, honesty and other qualities of investigators.

**Suitability:**

**This method is suitable when:**

1. Area of investigation is limited.

2. A very high degree of accuracy is required.

3. 3. The result of investigation is to be kept secret.

4. Area of investigation is homogeneous i.e. having same qualities.

Indirect Oral investigation.

## **2. Indirect Oral Investigation:**

Here, the investigator collects the data indirectly by interviewing persons who are known to be close to the original persons or the incidence. This method is adopted when the original persons are not found or found to be reluctant to provide the required information. In this method a list of questions is prepared, and the witnesses are invited and made to answer the questions. The answers are recorded by the investigators.

### **his method has its merits and demerits also:**

#### **Merits:**

1. It can cover a wide area.
2. It needs less time, energy and money.
3. Third party does not conceal the facts.
4. Intelligence, skill and tact of the investigator brings accuracy.

#### **ADVERTISEMENTS:**

5. It is simple and convenient.

#### **Demerits:**

1. The information obtained from the other persons may not be reliable.
2. The third parties may be biased.
3. The witness chosen may not be proper for the cause or not expert in this field.

### **This method is suitable when:**

1. The direct approach to sources does not exist.
2. The person cannot be relied upon
3. The area of investigation is large.
4. The information is needed, is kept secret from person.

## **3. Information through Local Correspondents:**

In this method, the investigator does not work but appoints local agents or enumerators in different parts of the area. These agents or enumerators are asked to collect information and transmit it to



the investigator. This method is often adopted by government, semi-government and bigger private institutions to collect the information for any purpose.

**Following are the merits and demerits of this method:**

**Merits:**

1. This method is cheap or economical.

There is economy of time and labour in this method.

3. This method can be used where the area of inquiry is wide.

4. Quality of data is better as enumerators are trained.

**Demerits:**

1. Data are not reliable which are collected by this method.

2. If high degree of accuracy is required this method is not adopted.

3. The data are not original if collected by this method.

4. Collected data lack uniformity because they are collected by different correspondents and they differ in quality and taste.

**Suitability:**

**This method is suitable when:**

(i) Rough and approximate estimates are required.

(ii) The information is wanted at regular intervals.

(iii) The field of inquiry is wide.

**Questionnaires through Mail:**

An important method for the collection of data used for large area is that of questionnaires. In this method, a list of questions is prepared relating to the problem under investigation, is printed and then sent out to the informants through post. It is requested that it may be returned to the investigator properly filled up. A covering letter is also sent with the questionnaire. A stamped self addressed envelope is also attached.

**Following are the merits and demerits of this method:**

**Merits:**

1. There is the economy of time, labour and wealth.

2. This method is used if the field of inquiry is wide.
3. There is uniformity in the data because informants are directly involved,

**Demerits:**

1. Information can be collected from the educated persons only.
2. Its response is uncertain.
3. Supplementary questions cannot be asked as questions are fixed.
4. If questionnaire is tough, it may not be responded.
5. There are chances of non-response due to indifference of attitude.

**Suitability:**

**This method is suitable when:**

- (i) The informant is educated.
- (ii) The area of coverage is very wide.

**Precautions:**

1. Questions should be in a good sequence.
2. Questions should not be personal, as informant will not want to answer.
3. A request letter must be attached.
4. It should be accompanied by a self addressed, stamped envelope to ensure reply.

**5. Schedules Sent through Enumerators:**

In this method, list of questions or schedules are sent to the informants through the enumerators. They read the questions to the informants and record their answers on the same schedules. At first, enumerator explains the aims and objectives of the enquiry and asks them for co-operation.

The difference between the mailed questionnaire method and this method is that in the former case the informants themselves record the answers while in this case, the enumerators record the information. This method is used by governments and research institutions like National Council of Applied Economic Research (NCAER), Federation of Indian Chamber of Commerce and Industries (FICCI) etc.

**Like the other methods, this method also has merits and demerits:**

**Merits:**

1. Information can be got even from uneducated persons.
2. This information is more reliable and correct.
3. It covers wide area.
4. It is unaffected by the personal bias of the investigators.
5. There are fewer chances of non-responses as enumerator's visits personally.

**Demerits:**

1. It is costly because enumerators have to be paid.
2. It is time consuming as every informant is visited.
3. It requires trained enumerators, which are not easily available.
4. The personal bias of enumerator may lead to wrong conclusions.
5. It can only be used by big organizations.

**Suitability:****This method is suitable when:**

1. Area to be covered is wide.
2. No extra question is needed to be asked.
3. Service of enumerators is cheap and easily available.

**Secondary Data.**

When the data are collected by someone else for a purpose other than the researcher's current project and has already undergone the statistical analysis is called as **Secondary Data**.

Sources of Secondary data are

- I. Published sources
- II. Unpublished sources.

## **CLASSIFICATION:**

The collected data, also known as raw data or ungrouped data are always in an unorganized form and need to be organized and presented in meaningful and readily comprehensible form in order to facilitate further statistical analysis. It is, therefore, essential for an investigator to condense a mass of data into more and more comprehensible and assimilable form. The process of grouping into different classes or subclasses according to some characteristics is known as classification, tabulation is concerned with the systematic arrangement and presentation of classified data. This classification is the first step in tabulation.

For Example, letters in the post office are classified according to their destinations viz., Delhi, Madurai, Bangalore, Mumbai etc.,

## **TYPES OF CLASSIFICATION:**

Statistical data are classified in respect of their characteristics. Broadly, there are four basic types of classification namely

- a) Chronological classification
- b) Geographical classification
- c) Qualitative classification
- d) Quantitative classification

### **a) Chronological classification:**

In chronological classification the collected data are arranged according to the order of time expressed in years, months, weeks, etc., The data is generally classified in ascending order of time. For example, the data related with population, sales of a firm, imports and exports of a country are always subjected to chronological classification.

#### **Example:**

The estimates of birth rates in India during 1970 – 76 are

<b>Year</b>	1970	1971	1972	1973	1974	1975
<b>Birth Rate</b>	36.8	36.9	36.6	34.6	34.5	35.2

### **b) Geographical classification:**

In this type of classification the data are classified according to geographical region or place. For instance, the production of paddy in different states in India, production of wheat in different countries etc.,

**Example:**

<b>Country</b>	America	China	Denmark	France	India
<b>Yield of wheat in (kg/acre)</b>	1925	893	225	439	862

### c) Qualitative classification:

In this type of classification data are classified on the basis of same attributes or quality like sex, literacy, religion, employment etc., Such attributes cannot be measured along with a scale. For example, if the population to be classified in respect to one attribute, say sex, then we can classify them into two namely that of males and females. Similarly, they can also be classified into 'employed' or 'unemployed' on the basis of another attribute 'employment'.

Thus when the classification is done with respect to one attribute, which is dichotomous in nature, two classes are formed, one possessing the attribute and the other not possessing the attribute. This type of classification is called simple or dichotomous classification. A simple classification may be shown as under

<b>Population</b>	
<b>Male</b>	<b>Female</b>

The classification, where two or more attributes are considered and several classes are formed, is called a manifold classification. For example, if we classify population simultaneously with respect to two attributes, e.g sex and employment, then population are first classified with respect to 'sex' into 'males' and 'females'. Each of these classes may then be further classified into 'employment' and 'unemployment' on the basis of attribute 'employment' and as such Population are classified into four classes namely.

- (i) Male employed
- (ii) Male unemployed
- (iii) Female employed
- (iv) Female unemployed

Still the classification may be further extended by considering other attributes like marital status etc. This can be explained by the following chart

<b>Population</b>			
<b>Male</b>		<b>Female</b>	
<b>Employed</b>	<b>Unemployed</b>	<b>Employed</b>	<b>Unemployed</b>

### d) Quantitative classification:

Quantitative classification refers to the classification of data according to some characteristics that can be measured such as height, weight, etc., For example the students of a college may be classified according to weight as given below.

<b>Weight (in lbs)</b>	<b>No of Students</b>
90-100	50
100-110	200
110-120	260
120-130	360
130-140	90
140-150	40
Total	1000

In this type of classification there are two elements, namely (i) the variable (i.e) the weight in the above example, and (ii) the frequency in the number of students in each class. There are 50 students having weights ranging from 90 to 100 lb, 200 students having weight ranging between 100 to 110 lb and so on.

### **TABULATION:**

Tabulation is the process of summarizing classified or grouped data in the form of a table so that it is easily understood and an investigator is quickly able to locate the desired information. A table is a systematic arrangement of classified data in columns and rows. Thus, a statistical table makes it possible for the investigator to present a huge mass of data in a detailed and orderly form. It facilitates comparison and often reveals certain patterns in data which are otherwise not obvious. Classification and 'Tabulation', as a matter of fact, are not two distinct processes. Actually they go together. Before tabulation data are classified and then displayed under different columns and rows of a table.

### **ADVANTAGES OF TABULATION:**

Statistical data arranged in a tabular form serve following objectives:

1. It simplifies complex data and the data presented are easily understood.
2. It facilitates comparison of related facts.

3. It facilitates computation of various statistical measures like averages, dispersion, correlation etc.
4. It presents facts in minimum possible space and unnecessary repetitions and explanations are avoided. Moreover, the needed information can be easily located.
5. Tabulated data are good for references and they make it easier to present the information in the form of graphs and diagrams.

### **PREPARING A TABLE:**

The making of a compact table itself an art. This should contain all the information needed within the smallest possible space. What the purpose of tabulation is and how the tabulated information is to be used are the main points to be kept in mind while preparing for a statistical table. An ideal table should consist of the following main parts:

1. Table number
2. Title of the table
3. Captions or column headings
4. Stubs or row designation
5. Body of the table
6. Footnotes
7. Sources of data

### **Table Number:**

A table should be numbered for easy reference and identification. This number, if possible, should be written in the center at the top of the table. Sometimes it is also written just before the title of the table.

### **Title:**

A good table should have a clearly worded, brief but unambiguous title explaining the nature of data contained in the table. It should also state arrangement of data and the period covered. The title should be placed centrally on the top of a table just below the table number (or just after the table number in the same line).



**Captions or column Headings:**

A caption in a table stands for brief and self explanatory headings of vertical columns. Captions may involve headings and sub-headings as well. The unit of data contained should also be given for each column. Usually, a relatively less important and shorter classification should be tabulated in the columns.

**Stubs or Row Designations:**

Stubs stands for brief and self explanatory headings of horizontal rows. Normally, a relatively more important classification is given in rows. Also a variable with a large number of classes is usually represented in rows. For example, rows may stand for score of classes and columns for data related to sex of students. In the process, there will be many rows for scores classes, but only two columns for male and female students.

**Body:**

The body of the table contains the numerical information on frequency of observations in the different cells. This arrangement of data is according to the description of captions and stubs.

**Footnotes:**

Footnotes are given at the foot of the table for an explanation of any fact or information included in the table which needs some explanation. Thus, they are meant for explaining or providing further details about the data, which have not been covered in title, captions and stubs.

**Sources of data:**

Lastly, one should also mention the source of information from which data are taken. This may preferably include the name of the author, volume, page and the year of publication. This should also state whether the data contained in the table is of 'primary or secondary' nature. **Date:**

The date of preparation of the table should be given.

**Unit of measurement:**

If the unit of measurement is uniform throughout the table, it is stated at the top right-hand corner of the table along with the title. If different rows and columns contain figures in different units, the units may be stated along with “stubs”, or, “captions”. Very large figures may be rounded up, but the method of rounding should be explained.